Enhancing Reproducibility and Transparency of Research Findings

107th Meeting of the Advisory Committee to the Director

December 5th, 2013

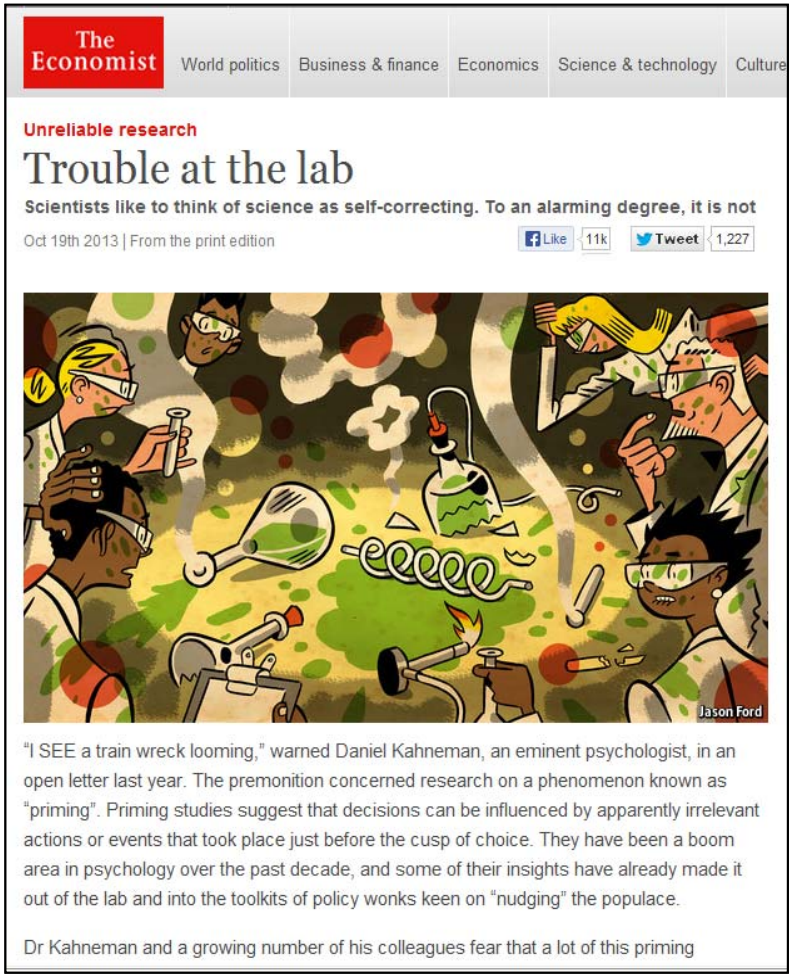
Lawrence A. Tabak, DDS, PhD

Principal Deputy Director, NIH

Department of Health and Human Services

Background

- Reproducibility and transparency of research findings have been noted as an issue in multiple publications.
  - This is a problem in all areas of research
  - This issue has been observed in both clinical and preclinical research, though NIH focus is preclinical research



The Economist World politics Business & finance Economics Science & technology Culture

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

Like 11k Tweet 1,227

Jason Ford

"I SEE a train wreck looming," warned Daniel Kahneman, an eminent psychologist, in an open letter last year. The premonition concerned research on a phenomenon known as "priming". Priming studies suggest that decisions can be influenced by apparently irrelevant actions or events that took place just before the cusp of choice. They have been a boom area in psychology over the past decade, and some of their insights have already made it out of the lab and into the toolkits of policy wonks keen on "nudging" the populace.

Dr Kahneman and a growing number of his colleagues fear that a lot of this priming

The Economist – October 19th

Beware the creeping cracks of bias

Evidence is mounting that research is riddled with systematic errors. Left unchecked, this could erode public trust, warns **Daniel Sarewitz**.

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

Statistical Design Considerations in Animal Studies Published Recently in *Cancer Research*

Kenneth R. Hess

Raise standards for preclinical cancer research

C. Glenn Begley and **Lee M. Ellis** propose how methods, publications and incentives must change if patients are to benefit.

Why animal research needs to improve

Many of the studies that use animals to model human diseases are too small and too prone to bias to be trusted, says **Malcolm Macleod**.

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research

David Moher*[1,2], Iveta Simera[3], Kenneth F Schulz[4], John Hoey[5] and Douglas G Altman[3]

Reforming Science: Methodological and Cultural Reforms

Drug targets slip-sliding away

The starting point for many drug discovery programs is a published report on a new drug target. Assessing the reliability of such papers requires a nuanced view of the process of scientific discovery and publication.

Translating animal research into clinical benefit

Poor methodological standards in animal studies mean that positive results may not translate to the clinical domain

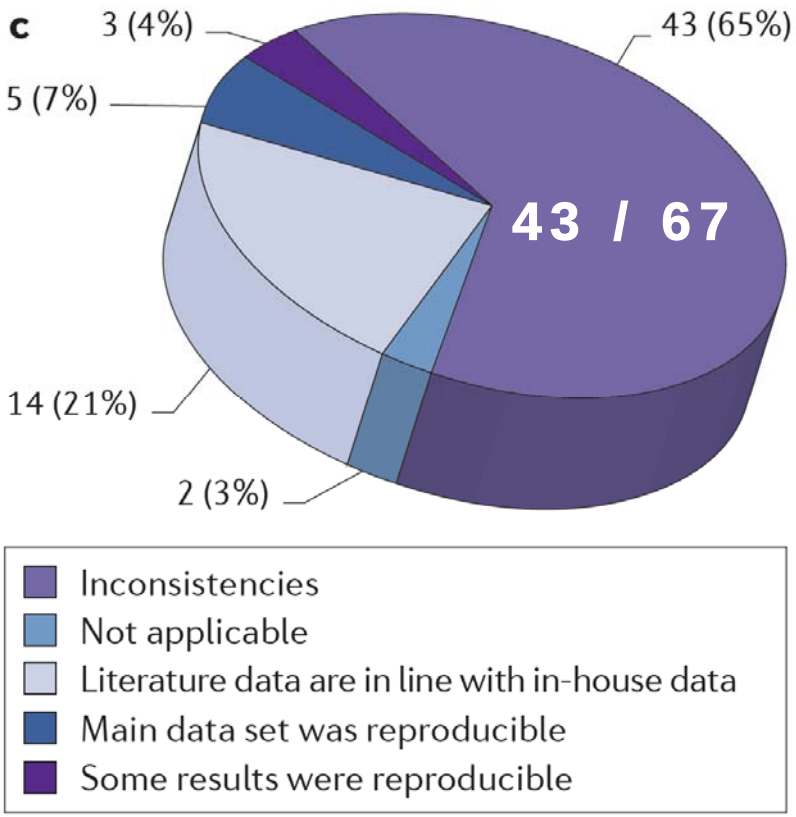
Courtesy of Dr. S. Silberberg, NINDS

Almost 2/3 of 67 in-house projects could not replicate data published by others

Believe it or not: how much can we rely on published data on potential drug targets?

Prinz, Schlange and Asadullah

Bayer HealthCare



Nature Reviews Drug Discovery, 2011; 10:712-713

Adapted from Dr. S. Silberberg, NINDS

ANALYSIS

doi:10.1038/nature12831

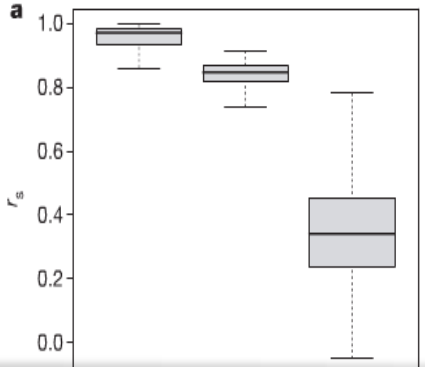
Inconsistency in large pharmacogenomic studies

Benjamin Haibe-Kains[1,2], Nehme El-Hachem[1], Nicolai Juul Birkbak[3], Andrew C. Jin[4], Andrew H. Beck[4]*, Hugo J. W. L. Aerts[5,6,7]* & John Quackenbush[5,8]*

Two large-scale pharmacogenomic studies were published recently in this journal. Genomic data are well correlated between studies; however, the measured drug response data are highly discordant. Although the source of inconsistencies remains uncertain, it has potential implications for using these outcome measures to assess gene–drug associations or select potential anticancer drugs on the basis of their reported results.

Patients with cancer often exhibit heterogeneous responses to anticancer treatments, and evidence indicates that response is determined in part by patient-specific alterations in the somatic cancer genome and changes in gene expression[1]. Cancer cell line studies have long been used to test the efficacy of therapeutic agents and to explore genomic factors associated with drug response[2,3]. A number of studies have searched for gene expression signatures predictive of response; however, most only tested a limited number of genes, a small panel of drugs, or assayed drug response in a small number of cell lines[2,4,5].

Results from two large-scale pharmacogenomic studies—the Cancer



Background (cont.)

- Relevant NIH workshops in 2012
  - NINDS: “Optimizing the Predictive Value of Preclinical Research”, summarized in 11 October 2012 issue of *Nature*
    (Held in June)
  - NCI: Reproducibility and data standards
    (Held in September and December)
- NIH Leadership discusses underlying causes and the development of “pilot” interventions in 2013

Possible causes in difficulties reproducing data

- Misconduct - Falsification, Fabrication, or Plagiarism
  - In 2011, the Office of Research Integrity*:
    - Received 240 allegations
    - Opened 12 as cases
  - Misconduct is one cause, but not the focus of this effort

*http://ori.hhs.gov/images/ddblock/ori_annual_report_2011.pdf

Possible causes in difficulties reproducing data

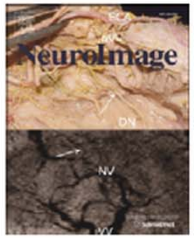
- ~~Misconduct - Falsification, Fabrication, or Plagiarism~~
- “Cartoon biology” – overemphasis on the “exciting, big picture” finding sometimes results in publications leaving out necessary details of experiments performed

NeuroImage 63 (2012) 289–300

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Full Length Articles

The secret lives of experiments: Methods reporting in the fMRI literature

Joshua Carp

University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, MI, 48109, USA

ARTICLE INFO

Article history:
Accepted 3 July 2012
Available online 10 July 2012

Keywords:
fMRI
Methods reporting
Reproducibility
Experimental design
Analysis methods
Statistical power

ABSTRACT

Replication of research findings is critical to the progress of scientific understanding. Accordingly, most scientific journals require authors to report experimental procedures in sufficient detail for independent researchers to replicate their work. To what extent do research reports in the functional neuroimaging literature live up to this standard? The present study evaluated methods reporting and methodological choices across 241 recent fMRI articles. Many studies did not report critical methodological details with regard to experimental design, data acquisition, and analysis. Further, many studies were underpowered to detect any but the largest statistical effects. Finally, data collection and analysis methods were highly flexible across studies, with nearly as many unique analysis pipelines as there were studies in the sample. Because the rate of false positive results is thought to increase with the flexibility of experimental designs, the field of functional neuroimaging may be particularly vulnerable to false positives. In sum, the present study documented significant gaps in methods reporting among fMRI studies. Improved methodological descriptions in research reports would yield significant benefits for the field.

Possible causes in difficulties reproducing data

- ~~Misconduct - Falsification, Fabrication, or Plagiarism~~
- "Cartoon biology" – overemphasis on the "exciting, big picture" finding sometimes results in publications leaving out necessary details of experiments performed
- Chance – Experiments performed correctly, but without appropriate replication
  - Difficulty in publication of "negative" findings

Possible causes in difficulties reproducing data

- ~~Misconduct - Falsification, Fabrication, or Plagiarism~~
- "Cartoon biology" – overemphasis on the "exciting, big picture" finding sometimes results in publications leaving out necessary details of experiments performed
- Chance – Experiments performed correctly, but without appropriate replication
  - Difficulty in publication of "negative" findings
- Poor experimental design – fundamental quality characteristics not reported/performed (e.g. blinded assessment, randomization, sample size calculations)

Insufficient reporting of methodological approaches is evident for pre-clinical studies

Table 3. Prevelence of selected quality characteristics in other experimental models

	Number of publications	Randomisation (%)	Blinded assessment of outcome (%)	Sample-size calculation (%)
Transgenic stroke studies	157	n/a	3	0
Stroke pathophysiology studies	166	5	18	0
Parkinson's disease	118	12	15	0
Multiple sclerosis	183	2	11	0

Trends Neurosci 2007; 30: 433-439

Adapted from Dr. S. Silberberg, NINDS

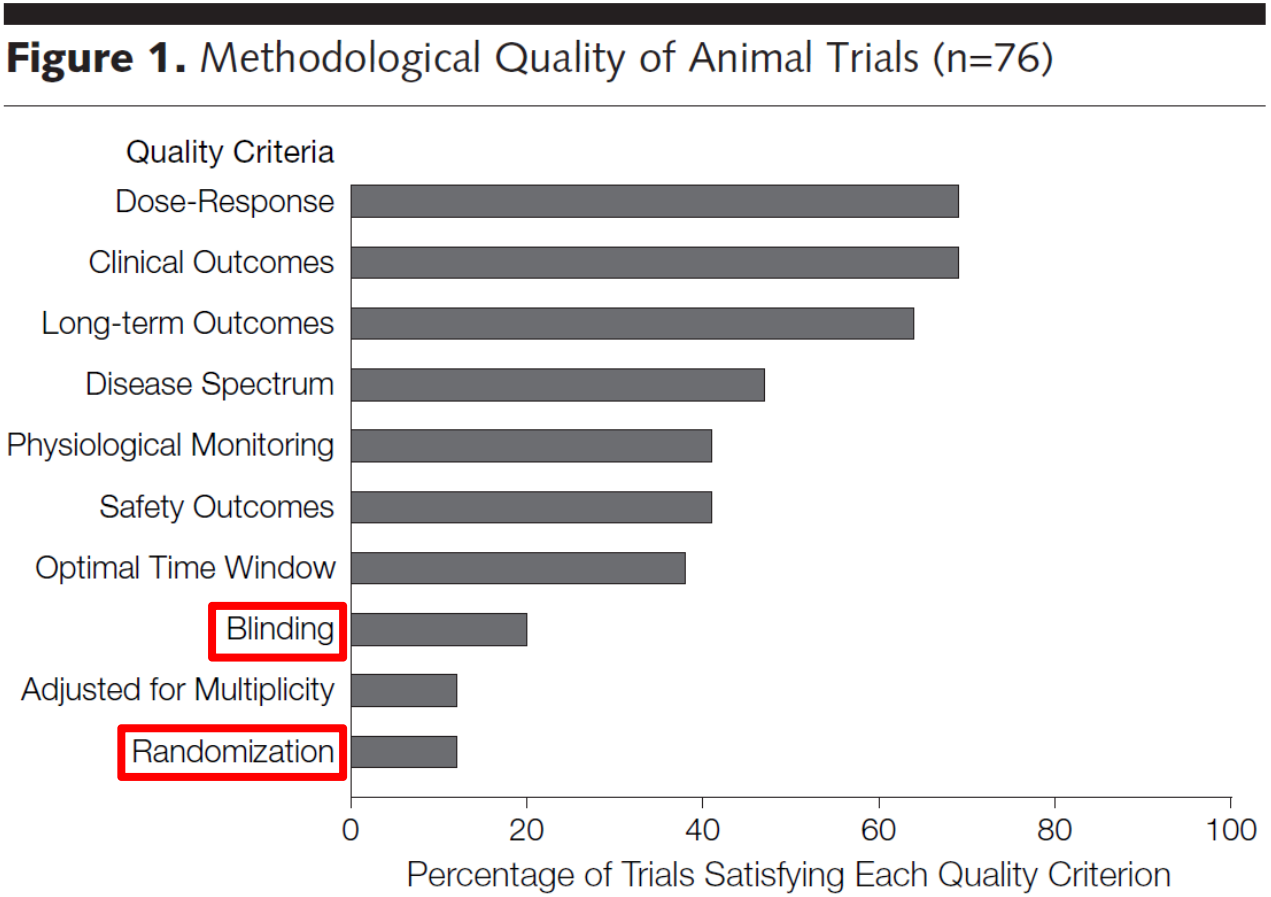
Deficient reporting is widespread

Journals:

- Cell
- Nature
- Science
- Nature Medicine
- Nature Genetics
- Nature Immunology
- Nature Biotechnology

>500 citations

Translated to human studies



Hackam and Redelmeier, *JAMA* 2006; 14: 1731-1732

Courtesy of Dr. S. Silberberg, NINDS

Possible causes in difficulties reproducing data

- ~~Misconduct - Falsification, Fabrication, or Plagiarism~~
- "Cartoon biology" – overemphasis on the "exciting, big picture" finding sometimes results in publications leaving out necessary details of experiments performed
- Chance – Experiments performed correctly, but without appropriate replication
  - Difficulty in publication of "negative" findings
- Poor experimental design – fundamental quality characteristics not reported/performed (e.g. blinded assessment, randomization, sample size calculations)
- Inadequate reporting of resources used

sites. The results of this experiment show that 54% of resources are not uniquely identifiable in publications, regardless of domain, journal impact factor, or reporting requirements. For example, in many cases the organism strain in which the experiment was performed or antibody that was used could not be identified. Our results show that identifiability is a serious problem for reproducibility. Based on these

repositories (such as model organism databases, and the Antibody Registry), as well as vendor sites. The results of this experiment show that 54% of resources are not uniquely identifiable in publications, regardless of domain, journal impact factor, or reporting requirements. For example, in many cases the organism strain in which the experiment was performed or antibody that was used could not be identified. Our results show that identifiability is a serious problem for reproducibility. Based on these results, we provide recommendations to authors, reviewers, journal editors, vendors, and publishers. Scientific efficiency and reproducibility depend upon a research-wide improvement of this substantial problem in science today.

COMMENT

Twenty tips for interpreting scientific claims

This list will help non-scientists to interrogate advisers and to grasp the limitations of evidence, say **William J. Sutherland, David Spiegelhalter** and **Mark A. Burgman.**

Calls for the closer integration of science in political decision-making have been commonplace for decades. However, there are serious problems in the application of science to policy — from energy to health and environment to education.

One suggestion to improve matters is to encourage more scientists to get involved in politics. Although laudable, it is unrealistic to expect substantially increased political involvement from scientists. Another proposal is to expand the role of chief scientific advisers[1], increasing their number, availability and participation in political processes. Neither approach deals with the core problem of scientific ignorance among many who vote in parliaments.

Perhaps we could teach science to politicians? It is an attractive idea, but which busy politician has sufficient time? In practice, policy-makers almost never read scientific papers or books. The research relevant to the topic of the day — for example, mitochondrial replacement, bovine tuberculosis or nuclear-waste disposal — is interpreted for them by advisers or external advocates. And there is rarely, if ever, a beautifully designed double-blind, randomized, replicated, controlled experiment with a large sample size and unambiguous conclusion that tackles the exact policy issue.

In this context, we suggest that the immediate priority is to improve policy-makers' understanding of the imperfect nature of science. The essential skills are to be able to intelligently interrogate experts and advisers, and to understand the quality, limitations and biases of evidence. We term these interpretive scientific skills. These skills are more accessible than those required to understand the fundamental science itself, and can form part of the broad skill set of most politicians.

To this end, we suggest 20 concepts that should be part of the education of civil servants, politicians, policy advisers and journalists — and anyone else who may have to interact with science or scientists. Politicians with a healthy scepticism of scientific advocates might simply prefer to arm themselves with this critical set of knowledge.

We are not so naive as to believe that improved policy decisions will automatically follow. We are fully aware that scientific judgement itself is value-laden, and that bias and context are integral to how data are collected and interpreted. What we offer is a simple list of ideas that could help decision-makers to parse how evidence can contribute to a decision, and potentially to avoid undue influence by those with vested interests. The harder part — the social acceptability of different policies — remains in the hands of politicians and the broader political process.

Of course, others will have slightly different lists. Our point is that a wider

1. Differences and chance cause variation
2. No measurement is exact
3. Bias is rife
4. Bigger is usually better for sample size
5. Correlation does not imply causation
6. Regression to the mean can mislead
7. Extrapolating beyond the data is risky
8. Beware the base-rate fallacy
9. Controls are important
10. Randomization avoids bias
11. Seek replication, not pseduoreplication
12. Scientists are humans
13. Significance is significant
14. Separate no effect from non-significance
15. Effect size matters
16. Study relevance limits generalization
17. Feelings influence risk perception
18. Dependencies change the risks
19. Data can be dredged or cherry picked
20. Extreme measurements may mislead

Underlying issues

- Poor training
- Poor evaluation
- Difficulty in publishing negative findings
- Perverse reward incentives

Learned Publishing, 24:95–97
doi:10.1087/20110203

~$30,000!

Table 1 Monetary reward system in Zhejiang University

Journal classification	Monetary award
Nature or *Science*	200,000 RMB (first author); decreased by 50% according to the sequence of authors
SCI journals (first author)	
IF < 1	2,000 RMB
1 ≤ IF < 3	3,000 RMB
3 ≤ IF < 5	4,000 RMB
5 ≤ IF < 10	5,000 RMB
IF ≥ 10	14,000 RMB
EI journals (first author)	1,800 RMB
ISTP (first author)	600 RMB

Principles for addressing the underlying issues

- Raise community awareness
- Enhance formal training
- Protect quality of funded and published research with a more systematic review process
- Address issues of pressure and stability for investigators

Trans-NIH actions

- NIH is discussing reproducibility and transparency of research findings with stakeholder communities to alert them to the issues and solicit feedback.
- Office of Intramural Research is creating and will pilot a new module on research integrity, as it relates to experimental biases and study design, to ethics training course required for NIH intramural fellows. This expected to be ready for testing in the Spring.
- Once tested, the Office of Extramural Research will make available on the web and encourage adoption (or equivalent) by extramural training programs for fellows and trainees.

Trans-NIH actions
Implementation of pilots

- NIH will implement pilots to address to key concerns:
  - Evaluate the “scientific premise” of grant applications
  - Develop a checklist to ensure more systematic evaluation of grant applications
  - Determine approaches needed to reduce “perverse incentives”, e.g.
    - Design changes to bio-sketch requirements
    - Longer-term support for investigators
  - Support replication studies

Trans-NIH actions
Implementation of pilots

- NIH will implement pilots to address to key concerns
- Important issues to consider as the pilots developed:
  - One size does not fit all
  - Effects on experienced vs. early-career researchers
  - Costs of additional data
  - Potential added burden to review process

PubMed Commons Blog

Keeping you up to date about PubMed Commons

Home About FAQs

PubMed Commons going public soon

Posted on November 26, 2013

It's been a month since the beta launch of PubMed Commons, the pilot system that enables authors' discussion and sharing of information about publications in PubMed.

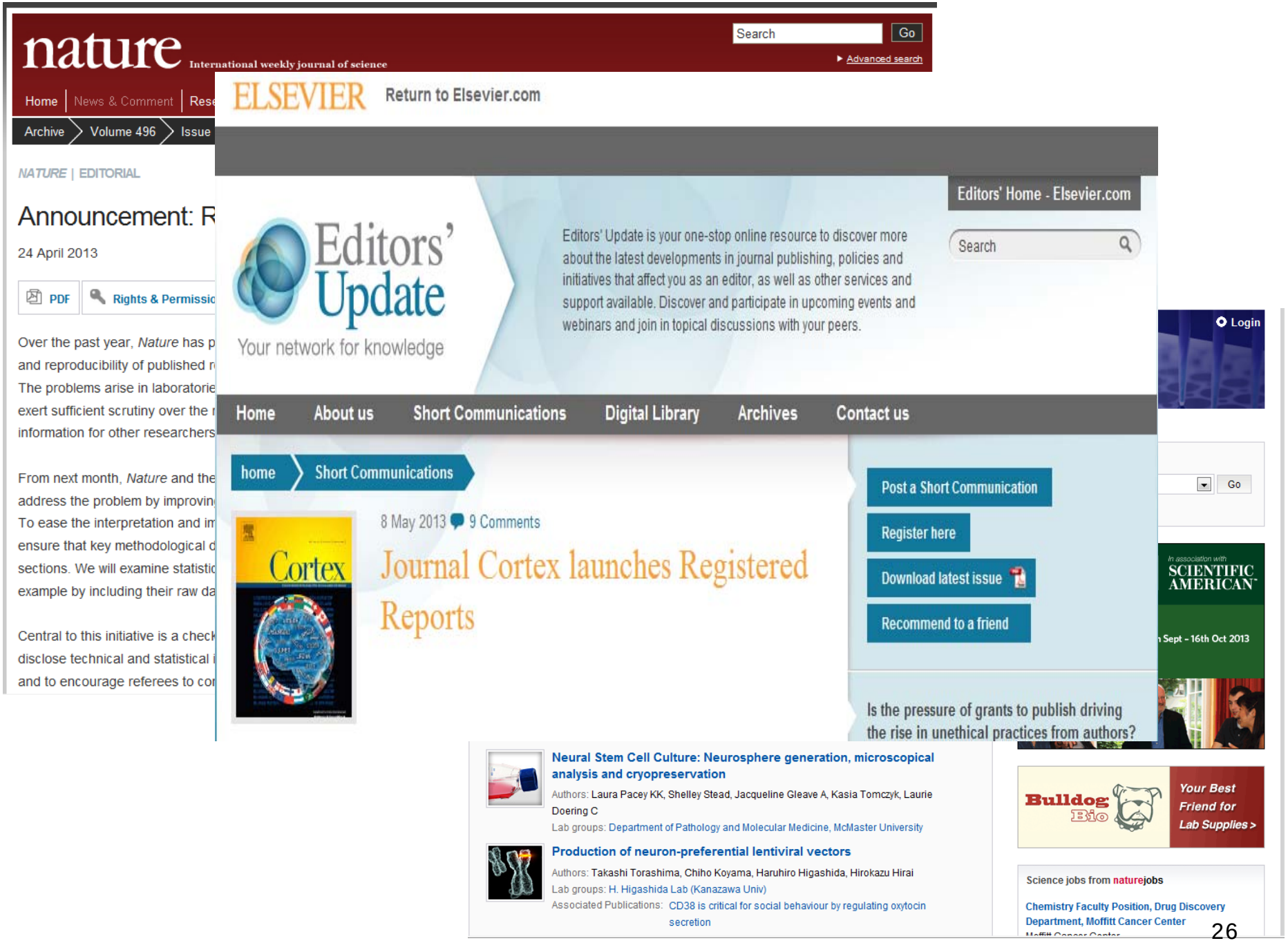
The first public version of the PubMed Commons pilot will be released in the coming weeks. All users of PubMed will be able to see and cite comments.

Trans-NIH actions (cont.)

- Convene meeting of Study Section Chairs, Board of Scientific Counselors (BSC) Chairs

Trans-NIH actions (cont.)

- Convene meeting of Study Section Chairs, Board of Scientific Counselors (BSC) Chairs
- Invite Journal Editors to meeting to discuss common opportunities


nature
International weekly journal of science
Search
Go
Advanced search
Home
News & Comment
Archive
Volume 496
Issue
NATURE | EDITORIAL
Announcement: R
24 April 2013
PDF
Rights & Permissic
Over the past year, Nature has p
and reproducibility of published r
The problems arise in laboratorie
exert sufficient scrutiny over the r
information for other researchers
From next month, Nature and the
address the problem by improvin
To ease the interpretation and im
ensure that key methodological d
sections. We will examine statistic
example by including their raw da
Central to this initiative is a check
disclose technical and statistical i
and to encourage referees to cor
ELSEVIER
Return to Elsevier.com
Editors' Update
Your network for knowledge
Editors' Update is your one-stop online resource to discover more about the latest developments in journal publishing, policies and initiatives that affect you as an editor, as well as other services and support available. Discover and participate in upcoming events and webinars and join in topical discussions with your peers.
Editors' Home - Elsevier.com
Search
Home
About us
Short Communications
Digital Library
Archives
Contact us
home
Short Communications
Cortex
8 May 2013
9 Comments
Journal Cortex launches Registered Reports
Post a Short Communication
Register here
Download latest issue
Recommend to a friend
Is the pressure of grants to publish driving the rise in unethical practices from authors?
Login
Go
In association with
SCIENTIFIC AMERICAN
Sept - 16th Oct 2013
Neural Stem Cell Culture: Neurosphere generation, microscopical analysis and cryopreservation
Authors: Laura Pacey KK, Shelley Stead, Jacqueline Gleave A, Kasia Tomczyk, Laurie Doering C
Lab groups: Department of Pathology and Molecular Medicine, McMaster University
Production of neuron-preferential lentiviral vectors
Authors: Takashi Torashima, Chiho Koyama, Haruhiro Higashida, Hirokazu Hirai
Lab groups: H. Higashida Lab (Kanazawa Univ)
Associated Publications: CD38 is critical for social behaviour by regulating oxytocin secretion
Bulldog Bio
Your Best Friend for Lab Supplies >
Science jobs from naturejobs
Chemistry Faculty Position, Drug Discovery Department, Moffitt Cancer Center

Elsevier Announces Article Retraction from Journal Food and Chemical Toxicology

"Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize," by Gilles Eric Séralini et al. has been retracted by the journal Food and Chemical Toxicology

Cambridge, MA, November 28, 2013

Elsevier announces that the article "Long term
Eric Séralini et al. has been retracted by the jo

The journal has issued the following retractio

The journal *Food and Chemical Toxicology* re
modified maize," which was published in this
of the published article and the data it reports,
making any public statements regarding this

Very shortly after the publication of this article,
described, the proper use of animals, and eve
paper. According to the journal's standard pra
response from the authors.[1] Due to the natu
review process and requested permission fro
made; however, it is in accordance with the jo
so requested.[2] The corresponding author a
wishes to acknowledge the co-operation of th
process.

Food and Chemical Toxicology 50 (2012) 4221–4231

Contents lists available at SciVerse ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize

Gilles-Eric Séralini [a,*], Emilie Clair [a], Robin Mesnage [a], Steeve Gress [a], Nicolas Defarge [a], Manuela Malatesta [b], Didier Hennequin [c], Joël Spiroux de Vendômois [a]

[a] University of Caen, Institute of Biology, CRIIGEN and Risk Pole, MRSH-CNRS, EA 2608, Esplanade de la Paix, Caen Cedex 14032, France
[b] University of Verona, Department of Neurological, Neuropsychological, Morphological and Motor Sciences, Verona 37134, Italy
[c] University of Caen, UR ABTE, EA 4651, Bd Maréchal Juin, Caen Cedex 14032, France

ARTICLE INFO

Article history:
Received 11 April 2012
Accepted 2 August 2012
Available online 19 September 2012

Keywords:
GMO
Roundup
NK603
Rat
Glyphosate-based herbicides
Endocrine disrupting effects

ABSTRACT

The health effects of a Roundup-tolerant genetically modified maize (from 11% in the diet), cultivated with or without Roundup, and Roundup alone (from 0.1 ppb in water), were studied 2 years in rats. In females, all treated groups died 2–3 times more than controls, and more rapidly. This difference was visible in 3 male groups fed GMOs. All results were hormone and sex dependent, and the pathological profiles were comparable. Females developed large mammary tumors almost always more often than and before controls, the pituitary was the second most disabled organ; the sex hormonal balance was modified by GMO and Roundup treatments. In treated males, liver congestions and necrosis were 2.5–5.5 times higher. This pathology was confirmed by optic and transmission electron microscopy. Marked and severe kidney nephropathies were also generally 1.3–2.3 greater. Males presented 4 times more large palpable tumors than controls which occurred up to 600 days earlier. Biochemistry data confirmed very significant kidney chronic deficiencies; for all treatments and both sexes, 76% of the altered parameters were kidney related. These results can be explained by the non linear endocrine-disrupting effects of Roundup, but also by the overexpression of the transgene in the GMO and its metabolic consequences.

Trans-NIH actions (cont.)

- Convene meeting of Study Section Chairs, Board of Scientific Counselors (BSC) Chairs
- Invite Journal Editors to meeting to discuss common opportunities
- Continue dialogue with stakeholders – professional societies, industry, academics, patient advocacy groups

Extramural Research Community

Reproducibility Initiative receive
landmark cancer studies

October 16, 2013 | Posted by Elizabeth in Science Exch

Home About Press Jobs Contact

Rewarding Reproducible Re

The COS advocates openness, integrity, and reproducibility of scientific research.

For Scientists

COS maintains the Open Science Framework to archive, share, register, and manage the research lifecycle. COS supports the Open Science Collaboration, develops tools to enable open practices, and provides grants for replications and other projects.

Read more

For Journals, Funders, and Societies

COS supports infrastructure for preregistration and open practices, produces badges to incentivize open practices, maintains back-end tools for journals, and provides grants for open innovations.

Read more

For Developers

COS supports open source projects for science, connects via API with other services, hires open source developers, and provides grants for new open projects.

Read more

nature International weekly journal of science

Search Go

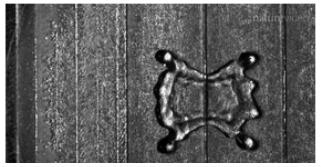
Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

News & Comment › News › 2013 › November › Article

NATURE | NEWS

Top Story



Psychologists strike a blow for reproducibility

Thirty-six labs collaborate to check 13 earlier findings.

Ed Yong

26 November 2013

Rights & Permissions

A large international group set up to test th
reproduced the results of 10 out of 13 past
could not be reproduced.

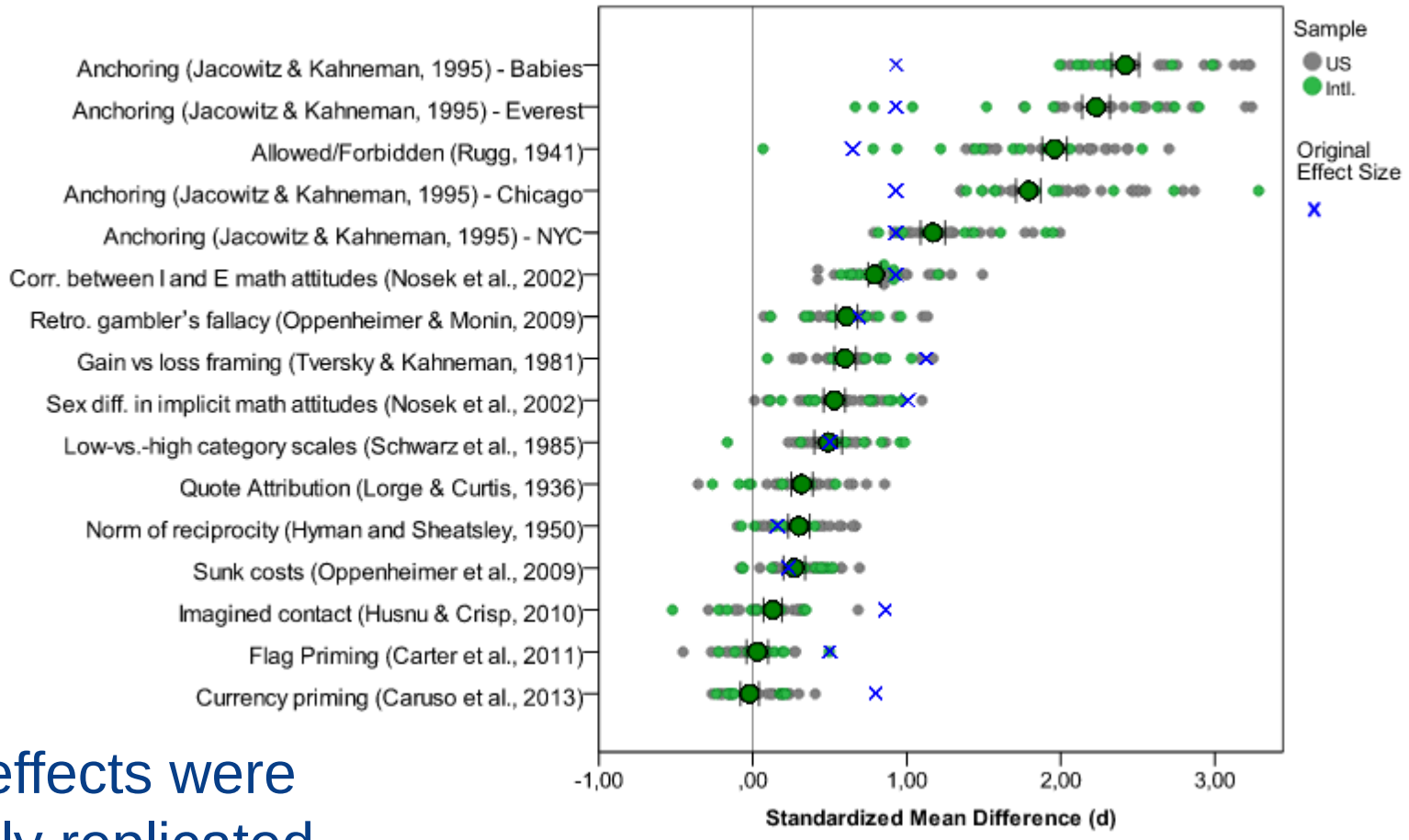
Psychology has been buffeted in recent ye
after repeated failures to replicate classic s
study was flawed, the new experiment was
settings or groups of people.

To tackle this 'replicability crisis', 36 resear
Labs Replication Project to repeat 13 psyc
consortium combined tests from earlier exp
questionnaire — meant to take 15 minutes
it to 6,344 volunteers from 12 countries.

The team chose a mix of effects that repre
psychological science, from classic experin
repeatedly replicated to contemporary one

Ten of the effects were consistently replica
These included classic results from econor

Many Labs Replication Project repeated 13 psychological studies using 6,344 volunteers from 12 countries



10 of the effects were consistently replicated

Complementary NIH efforts

- Ongoing projects separate from and/or complementary to the proposed pilots
  - NIH has and continues to collaborate with the Association for Psychological Science (APS) and the American Psychological Association (APA) on new and enhanced journal reporting standards (e.g., expanded Methods sections, addition of statistical sections).
  - NIA: Supports the Interventions Testing Program, where preclinical studies are conducted with multi-site duplication, rigorous methodology and statistical analysis.

Complementary NIH efforts (cont.)

- Ongoing projects separate from and/or complementary to the proposed pilots
  - NHGRI: Expectations of validation studies are an inherent part of the review of functional genomics studies and bioinformatics tool development.
  - NIDDK: Supports Mouse Metabolic Phenotyping Centers, which provide the scientific community with standardized, high-quality phenotyping services.
  - NINDS: Established a Scientific Rigor Working Group to forge action plans for rigor-focused efforts.




NIH...

Lawrence.Tabak@nih.gov

Turning Discovery Into Health